

Extracting work from quantum systems

Paul Skrzypczyk,¹ Anthony J. Short,² and Sandu Popescu²

¹*Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, United Kingdom*

²*H. H. Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol, BS8 1TL, United Kingdom*

We consider the task of extracting work from quantum systems in the resource theory perspective of thermodynamics, where free states are arbitrary thermal states, and allowed operations are energy conserving unitary transformations. Taking as our work storage system a ‘weight’ we prove the second law and then present simple protocols which extract *average* work equal to the free energy change of the system – the same amount as in classical thermodynamics. Crucially, for systems in ‘classical’ states (mixtures of energy eigenstates) our protocol works on a single copy of the system. This is in sharp contrast to previous results, which showed that in case of *almost-deterministic* work extraction, collective actions on multiple copies are necessary to extract the free energy. This establishes the fact that free energy is a meaningful notion even for individual systems in classical states. However, for non-classical states, where coherences between energy levels exist, we prove that collective actions are necessary, so long as no external sources of coherence are used.

I. INTRODUCTION

Thermodynamics forms part of the bedrock of our current understanding of the physical world. It has remained unchanged despite huge revolutions in physics, such as relativity and quantum theory, and few believe it will ever fail. Over time, it has been applied to situations well outside its original domain; from black holes [1, 2], to quantum engines comprised of only a few qubits [3–6]. A fundamental question is that of the applicability of thermodynamics to quantum systems [7, 8]; it is this question that we wish to address in this paper.

One of the most fundamental questions one would like to ask is how much work can be extracted from a system, given access to a thermal bath at temperature T . This question, for generic quantum systems and Hamiltonians has been investigated intensively, e.g. in [9–15] yet, until very recently only partial results have been obtained and it remained an open problem. Very recently however strong general results were obtained [16], made possible by employing a new conceptual framework, namely that of “resource theories”.

The advent of resource theories is one of the main paradigm shifts which occurred within quantum information theory; the main idea is to bring to the forefront the study of *resources* needed to implement various tasks. The archetypal resource theory in this context is entanglement theory [17, 18]. However, numerous other examples have now been developed, including reference frames [19], purity [20] and asymmetry [21], which has culminated in a general formalised notion of a resource theory [22]. In [23] the authors introduced an appealing resource theory of thermodynamics, based on the class of allowed operations first studied in [24].

Here we revisit and modify the resource theory framework of thermodynamics as originally put forward in [23]. In this new context we achieve two main results: First we prove the second law of thermodynamics. Second, as described in more detail below, we show that free energy is a relevant quantity for individual quantum systems.

The latter result is particularly relevant following the re-

sults of [16] which call into question the role of free energy for individual quantum systems. Classical thermodynamics tells us that the total amount of work we are able to extract from a system is given by its change in *free energy*, which was also supported by previous partial quantum results [9–15]. Yet in [16] it was shown that work equal to free energy can be extracted only if we *collectively* process many copies of the same system. When acting on each copy individually, the amount of work that can be extracted is generally significantly less than the free energy. These results therefore suggest that the free energy is not the relevant quantity for individual systems.

Here we revisit the issue of work extraction and show that free energy is a significant quantity for individual systems. Our paradigm is similar to that of [16] but differs in an essential aspect. In [16] they considered *almost deterministic*¹ work extraction, from the ‘single-shot’ viewpoint which has received much attention lately [14, 25–28]. Here, in contrast, we will consider *average* work extraction. We prove that there exist simple protocols which extract *average* work equal to the free energy change of the system. Crucially, for systems in ‘classical’ states (mixtures of energy eigenstates) our protocol works on a single copy of the system. This establishes the fact that free energy is a meaningful notion even for individual systems, when their states are classical.

However, for non-classical states, where coherences between energy levels exist, we prove that collective actions are necessary, so long as no external sources of coherence are used. More precisely, we prove that when acting on an individual system the amount of average work that can be extracted is equal to the free energy of its decohered state – the free energy due to coherences is ‘locked’ and cannot be used. As more copies are collectively processed, some of this locked free energy is released – we can extract the free energy of the decohered state of n copies, which is larger than n times the free energy of a single decohered system. In the

¹ Specifically, they require that a deterministic amount of work is extracted, except with some small probability ϵ of failure

limit of infinitely many copies, the average work extracted per copy reaches its entire free energy. Whether or not external coherent resources allow extraction of work equal to the free energy for individual systems, and how to account for such resources (given that they may provide free energy themselves even without the system) remains an open question.

The paper is organised as follows. In Sec. II we go through the preliminaries, including a brief description of the resource theory of thermodynamics in Sec. II A, work storage systems in Sec. II B and a proof that the second law holds in Sec. III. In Sec. IV we study work extraction from diagonal systems, beginning with a motivating example in Sec. IV A, before moving onto a full qubit protocol in Sec. IV B. We discuss a special case in Sec. IV C, present a full protocol for arbitrary dimension systems in Sec. IV D and prove optimality in Sec. IV E. In Sec. V we show how work can be extracted from coherence by acting collectively on multiple copies of the system. We finish by briefly discussing how the results of this paper can be understood from the perspective of ‘virtual temperatures’ in Sec. VI, before concluding in Sec. VII.

II. PRELIMINARIES

A. The resource theory of thermodynamics

The first ingredient in a resource theory is a set of free resources. By free we mean that we have an unlimited supply of this resource, since it is deemed easy to acquire. In the present context of thermodynamics this free resource is a *thermal bath at temperature T* . More precisely, we assume that we have an unlimited supply of finite-dimensional systems, with any desired Hamiltonian H , in their thermal state $\tau_\beta(H)$,

$$\tau_\beta(H) = \frac{1}{\mathcal{Z}} e^{-\beta H} \quad (1)$$

where $\beta = 1/T$ is the inverse temperature², and $\mathcal{Z} = \text{tr}(e^{-\beta H})$ is the partition function.

The second ingredient is a set of allowed operations, those which one has the ability to perform with ease. In the original resource framework these allowed operations were *any energy conserving unitary transformation*. In particular, any unitary operation which commutes with the total (non-interacting) Hamiltonian of all the systems under consideration. This allows one to transform systems in a controlled way, but without admitting an external source of energy.

However, we will also introduce an additional element into the resource framework, a *work storage system*, whose sole purpose is to provide and store work for thermodynamic transformations. In the next section, we will place some additional constraints on the allowed operations,

which prevent us from ‘cheating’ by using work storage system for other purposes (e.g. by using it as a cold reservoir in a heat engine), and allow us to use the same work storage system for multiple thermodynamic protocols.

Given these ingredients, any additional system which is not in thermal equilibrium with a bath at temperature T is a useful resource. The question that remains is to quantify how useful different resource states are.

B. Work storage systems

One way to quantify the usefulness of different resource states is to find a transformation from every resource state to some *fiducial* resource which allows for comparison in a well-defined and easy manner. In the resource theory of entanglement, for example, one can quantify the entanglement of a state by the optimal asymptotic number of maximally entangled states that can be created per resource state, or in other words the optimal rate at which states can be converted into maximally entangled states. Different states will have different optimal rates, and thus the rate quantifies how useful they are. We seek here an analogous fiducial resource for thermodynamics.

Work is the desired output in this case, and different states can produce work at different rates. Because we wish to study a closed quantum system, with conserved total energy, we desire a system where we can store the work extracted from resource states. Similarly to entanglement theory, where the maximally entangled state is chosen for convenience, here we wish to find a work storage system which we deem to be most convenient.

In previous work [16] a candidate was put forward – raising a qubit deterministically from its ground state to its excited state. This qubit was termed a *wit*, short for *work bit*. However, choosing the energy gap of the work bit requires advance knowledge of the work to be extracted, and this model does not translate well to non-deterministic work extraction, which we will be interested in here. Furthermore, we would prefer to be able to use a single work storage system as a ‘battery’ capable of gaining and expending work in multiple thermodynamic processes.

The alternative work storage system we will consider here is a suspended *weight*, which is raised or lowered when work is done on or by it. In particular, we will consider a quantum system whose height is given by the position operator \hat{x} , with Hamiltonian $H_w = mg\hat{x}$ representing its gravitational potential energy³. For simplicity, we choose $mg = 1 \text{ J/m}$, such that the value of \hat{x} directly denotes the work stored by the mass. Such a system has a long history of being used as a work storage system in classical thermodynamics [29].

³ Note that it is not strictly necessary for the height of the weight to be a continuous degree of freedom. In Appendix G we show that the weight can be replaced by a system with discrete equally-spaced energy levels. This introduces a small error in the results, but this can be made as small as desired by appropriately scaling the spacing of the weight.

² We set $k_B = 1$ throughout for convenience

We will place two additional constraints on the allowed dynamics governing interactions with the weight:

- (i) The average amount of work extracted in an allowed protocol must be independent of the initial state of the weight.
- (ii) All allowed unitaries should commute with translation operations on the weight. This reflects the translational symmetry of the weight system, and the fact that only displacements in its height are important.

These constraints are intended to prevent us from ‘cheating’ by using the weight for purposes other than as a work storage system (e.g. as a cold reservoir, or a source of coherence). For example if the weight starts in a pure state, it has zero entropy, so one needs to be careful to avoid unwittingly using this as a source of free energy, in addition to the system. Constraint (i) encapsulates this requirement.

Constraint (ii) is a bit more subtle. On the one hand, it is very plausible that it is a consequence of constraint (i). We however have not been able to show this. On the other hand, even if it does not follow from (i), we would still like to assume it, as it seems wrong to use properties of the weight in the protocol, even if they do not affect the work extracted.

Furthermore, constraint (i) ensures that we can use the same work storage system for multiple thermodynamics protocols (or on several copies of the same state) without having to worry how the initial state has been modified by earlier procedures.

As we are free to choose any initial state of the weight due to constraint (i) above, for simplicity we will usually take the initial state of the weight to be a narrow normalised wavepacket $|0\rangle_w$ centred on the origin (i.e. $\langle 0|\hat{x}|0\rangle = 0$). When the same wavepacket is raised above the origin by a distance a , we will denote the state by $|a\rangle_w = \Gamma_a|0\rangle_w$, where Γ_a is the translation operator⁴.

III. THE SECOND LAW

We now show that the second law of thermodynamics holds in our framework, by proving a probabilistic version of the second law – that there is no protocol which extracts a positive quantity of average work from a thermal bath (i.e. that there is no way on average of turning heat into work) [30]. To show this we will use proof by contradiction.

Consider a thermal bath at temperature T and the weight, i.e. there is no additional system out of thermal equilibrium.

Let us first consider the energy changes during the protocol. Suppose that we are able to extract work from the bath and store it in the weight, $\Delta E_W > 0$. The average energy of the thermal bath must change by $\Delta E_B = -\Delta E_W$ due to energy conservation.

Now consider the entropy changes during the same protocol (in particular, the changes in von Neumann entropy $S(\rho) = -\text{tr}(\rho \log \rho)$). As the bath and weight are initially uncorrelated, their initial entropy is simply the sum of their individual entropies. Unitary transformations conserve the total entropy, $\Delta S_{BW} = 0$. However, as correlations can form during the protocol, the sum of the final entropies of the bath and weight can be greater than the sum of their initial entropies (as the entropy is subadditive). This means that

$$\Delta S_B + \Delta S_W \geq \Delta S_{BW} = 0 \quad (2)$$

Given an initial thermal state for the bath (with positive temperature), any change of the state which reduces its average energy must also reduce its entropy⁵, $\Delta S_B < 0$. However, within our framework all allowed protocols are such that the average work extracted is independent of the initial state of the weight; we are therefore free to choose any initial state of the weight we like. Consider now that the initial state of the weight is a very wide wavepacket. We show in Appendix B that the entropy change of the weight in this case can be made as small as desired, in particular we can make $\Delta S_W < |\Delta S_B|$. This would result in violating (2). Hence there is a contradiction, and thus there is no way to extract work from the bath.

IV. EXTRACTING WORK FROM DIAGONAL SYSTEMS

We will begin our exploration of work extraction by considering the special case of extracting work from systems whose states are diagonal in the energy eigenbasis. This represents systems which are essentially classical, since coherence plays no role, and states are simply mixtures of energy eigenstates. We will present a protocol which extracts work equal to the change in free energy of the system. In contrast to existing protocols, here our protocol will take one copy of the system, and extract average work equal to the free energy of the system.

The protocol contains at its core a single elementary building block which is able to extract a vanishing amount of work blue, with maximal efficiency. The full protocol consists of many repetitions of this one step. We shall present first this building block in a motivating example and then show precisely how it is used to build the full protocol. In the main text the emphasis will be on the results of the protocol, with the more detailed analysis provided in the appendices.

⁴ The translation operator is $\Gamma_a = \exp(-ia\hat{p}/\hbar)$, where \hat{p} is the usual momentum operator satisfying $[\hat{x}, \hat{p}] = i\hbar$.

⁵ This follows from the fact that the thermal state is the maximal entropy state with given average energy. For more details see Appendix B

A. A motivating example

Let us consider the task of performing the following transformation on a single qubit

$$\begin{aligned} \rho_S &= (1-p)|0\rangle_S\langle 0| + p|1\rangle_S\langle 1| \\ \downarrow \\ \rho'_S &= (1-p+\delta p)|0\rangle_S\langle 0| + (p-\delta p)|1\rangle_S\langle 1| \end{aligned} \quad (3)$$

whose Hamiltonian is $H_S = E_S|1\rangle_S\langle 1|$ and where $\delta p \ll 1$. That is, we wish to consider the small transformation from a diagonal state to a nearby diagonal state. The initial and final state differ in free energy, and hence the goal is to extract an amount of work as close as possible to this difference, namely

$$\begin{aligned} \delta F_\beta &= F_\beta(\rho) - F_\beta(\rho') \\ &= (\langle E \rangle_\rho - TS(\rho)) - (\langle E \rangle_{\rho'} - TS(\rho')) \\ &\simeq \delta p (E_S - TS'_c(p)) \end{aligned} \quad (4)$$

where $F_\beta(\sigma) = \langle E \rangle_\sigma - TS(\sigma)$ is the free energy of σ with respect to a bath at inverse temperature β , $\langle E \rangle_\sigma = \text{tr}(H_S \sigma)$ is the average energy, $S_c(q) = -q \log q - (1-q) \log(1-q)$ is the (classical) binary entropy, $S'_c(q) = \frac{dS_c}{dq}$, and we have taken the first order expansion in δp to arrive at the final line. We shall now see that to first order we can extract exactly this amount of work from the system.

To do so let us bring in a thermal state from the bath – the free resource at our disposal – and choose its Hamiltonian such that the population in the excited state is $p - \delta p$. For thermal qubits the relation between excited state population r and energy spacing E_B is

$$E_B = T \log \left(\frac{1-r}{r} \right) = TS'_c(r) \quad (5)$$

thus we will choose the Hamiltonian of the bath qubit to be $H_B = E_B|1\rangle_B\langle 1|$ with

$$E_B = TS'_c(p - \delta p) \simeq TS'_c(p) - \delta p TS''_c(p) \quad (6)$$

Let us take the weight to be prepared initially in the state $|0\rangle_w$ of zero average energy, and consider applying an energy conserving unitary transformation U which interchanges

$$|0\rangle_B |1\rangle_S |x\rangle_w \leftrightarrow |1\rangle_B |0\rangle_S |x + E_S - E_B\rangle_w \quad (7)$$

for any x , whilst leaving all orthogonal states unchanged⁶.

⁶ U can be expressed in terms of the translation operator Γ_a as

$$U = |10\rangle\langle 01| \otimes \Gamma_\Delta + |01\rangle\langle 10| \otimes \Gamma_{-\Delta} + (|00\rangle\langle 00| + |11\rangle\langle 11|) \otimes \mathbb{1},$$

where $\Delta = E_S - E_B$. Note that this commutes with the free Hamiltonian $H_S + H_B + H_w$.

It is the change in average energy of the weight which is of interest. A straightforward calculation (given for completeness in the Appendix) shows that this is given by

$$\delta \langle E \rangle_w = \delta W = \delta p (E_S - E_B) \simeq \delta p (E_S - TS'_c(p)) \quad (8)$$

which is equal to δF_β to first order in δp .

Thus we observe that if we wish to make a change of state from ρ to ρ' , a way to proceed is to bring in a qubit from the bath in the state $\tau_\beta(H_B) = \rho'$. Whenever ρ is not the thermal state (given its Hamiltonian H_S), we find that $H_B \neq H_S$, and therefore there is a difference in energy between the excited states of the two qubits. It is thus not possible to swap the state of the two qubits using an energy conserving unitary. However, by bringing in the weight we can perform the unitary (7) which is energy conserving, and, at the level of the reduced states of the system and bath, indeed performs a swap. The system therefore ends up in the desired state ρ' , and we find that the change in average energy of the weight, when the state of the system and bath are close, approaches exactly the change in free energy of the system.

B. Full qubit protocol

The above insight can be used to build a full protocol which is able to extract an amount of work approaching the free energy change in transforming a system from a state ρ to its thermal state $\tau_\beta(H_S)$.

The basic idea is that, since in making a small change of state we have seen that the work extracted approaches the free energy change, we can divide a large change of state into many small changes and at each stage extract the desired amount of work.

Therefore, let us fix an integer $N \gg 1$, and consider N thermal qubits $\tau_\beta^{(k)}(H_B^{(k)})$, $k = 1, \dots, N$,

$$\begin{aligned} \tau_\beta^{(k)}(H_B^{(k)}) &= (1-r^{(k)})|0\rangle_{B_k}\langle 0| + r^{(k)}|1\rangle_{B_k}\langle 1| \\ r^{(k)} &= p + \frac{k}{N}(p_{eq} - p) \end{aligned} \quad (9)$$

where $p_{eq} = \frac{e^{-\beta E_S}}{1+e^{-\beta E_S}}$ is the excited-state probability in the thermal state $\tau_\beta(H_S)$. This collection of bath qubits have excited-state probabilities which vary linearly from that of the initial state of the system ρ_S to that of the thermal state of the system $\tau_\beta(H_S)$, in increments of $\frac{(p_{eq}-p)}{N}$. We will apply, in turn, the sequence of unitary transformations $\{U^{(k)}\}$, where $U^{(k)}$ acts on bath qubit k , system and weight and swaps the states

$$|0\rangle_{B_k} |1\rangle_S |x\rangle_w \leftrightarrow |1\rangle_{B_k} |0\rangle_S |x + E_S - E_B^{(k)}\rangle_w \quad (10)$$

whilst leaving all other orthogonal states unchanged.

Assuming once again that the weight starts in the state $|0\rangle_w$, it is the final state, and in particular the final average

energy of the weight which is of interest, as the number of steps N tends to infinity. In Appendix D we calculate explicitly this state in the asymptotic limit. The result is that the final reduced state in the infinite limit, denoted Ω_w^∞ , is

$$\Omega_w^\infty = (1-p) \left| T \log \left(\frac{1-p}{1-p_{eq}} \right) \right\rangle_w \left\langle T \log \left(\frac{1-p}{1-p_{eq}} \right) \right| + p \left| T \log \left(\frac{p}{p_{eq}} \right) \right\rangle_w \left\langle T \log \left(\frac{p}{p_{eq}} \right) \right| \quad (11)$$

The average energy is thus

$$\begin{aligned} \langle E \rangle_w^\infty &= T \left(p \log \left(\frac{p}{p_{eq}} \right) + (1-p) \log \left(\frac{1-p}{1-p_{eq}} \right) \right) \\ &= T D_c(p \| p_{eq}) = F_\beta(\rho) - F_\beta(\tau_\beta(H_S)) \end{aligned} \quad (12)$$

where $D_c(p \| p_{eq})$ is the relative binary entropy.

We observe that the average energy of the weight at the end of the protocol, and hence the average work extracted, approaches precisely the free energy difference in the asymptotic limit of infinitely many thermal qubits. Although this is promising, it should be noted that the final distribution of the weight does not have the form one may intuitively have asked for (i.e. a peak around a single value, at the free energy itself). Instead we find that the solution has perfect correlation between the initial state of the system and the final energy of the weight. It is worth noting that this interesting distribution coincides with the results found in [14] where, even though the details of the allowed transformations differ, it was shown that for classical systems, optimal work extraction protocols converge on this distribution in probability.

Note also that in the large N limit, although many thermal qubits are required, each of them is only slightly perturbed by the process.

Finally, we consider the case of n independent copies of the system – i.e. the macroscopic limit. To do so we simply repeat the protocol above on each qubit separately, using a single weight to store the work. In each case we have seen that the weight is transformed from a single peak (at $|0\rangle_w$ above) to two peaks. Since the protocol is linear, it follows immediately that the final distribution of the weight will be the n -fold convolution of the above distribution for a single qubit. From the central limit theorem one concludes that the final average energy of the weight is

$$\langle E \rangle_w = n (F_\beta(\rho) - F_\beta(\tau_\beta(H_S))) \quad (13)$$

with fluctuation of the order \sqrt{n} . For large n these fluctuations will be insignificant compared to the average, and we obtain an almost deterministic quantity of work, with the amount of work extracted per copy of the system equal to the free energy difference.

The results presented here are completely consistent with those presented in [23]. It should be noted however the stark difference in which the two protocols achieve the same goal. In [23] the analysis first considers all n systems together and uses ideas from information theory to compress information within large typical subspaces. Here, on the other hand, each

qubit is manipulated separately and nowhere is the idea of a typical subspace ever introduced, nevertheless they both achieve the same end result.

C. Special case: isothermal expansion

The protocol outlined in the previous subsection is the general protocol which allows for the extraction of work from an arbitrary diagonal state of a qubit, for any given Hamiltonian. There is one particular choice of state that will be particularly important in later sections. As such, we shall briefly look at it in more detail.

Consider a qubit for which the Hamiltonian vanishes ($E_S = 0$) and for which all the probability is initially in the state $|0\rangle_S$ ($p = 0$). Since the Hamiltonian vanishes, the system lives in a degenerate space, and since it is pure we have maximal information about the state of the system. This information can be used to extract work from the heat bath. Since all states of the system have equal energy, it is clear that the system provides only information, whilst it is the bath which provides the energy.

For a degenerate system the equilibrium state $\tau_\beta(0) = \frac{1}{2} \mathbb{1}$, thus after applying the protocol from the previous section, in the asymptotic limit of infinitely many bath qubits, the final state of the weight is given by

$$\Omega_w^\infty = |T \log 2\rangle_w \langle T \log 2| \quad (14)$$

corresponding to the extraction of an amount of work $W = T \log 2$ from the system. This is an analogous process to the isothermal expansion of a single-molecule gas from one side of a box into the full volume, as discussed in Szilard's well-known treatment of Maxwell's demon [31, 32].

It should be noted that contrary to the previous subsection, where the final state of the weight was found to consist of two peaks, in this special case of a pure initial state it consists of only a single peak. Thus it is not necessary to go to the many-copy regime to extract a deterministic amount of work in this case.

D. Extracting work from diagonal quantum states of arbitrary dimension

To conclude the analysis of diagonal systems it is now necessary to move beyond qubits, to quantum systems of arbitrary finite dimension d , and arbitrary Hamiltonians.

Our strategy is to connect a pair of energy levels in the system to a sequence of thermal qubits in the bath, shifting the occupation probability between these levels as in the qubit case (Sec. IV B). We then repeat this process for different pairs of levels until we have the desired final state. As the amount of work extracted in each step does not depend on the initial state of the weight (as proven in Appendix A), we can consider each step separately and simply sum the total work extracted.

An important difference from the qubit case is that the occupation probability of the two levels in the system that are coupled to the bath at any given stage need not sum to unity. For example consider a qutrit (or larger system) in the state

$$\rho_S = (1 - p - q)|0\rangle_S\langle 0| + q|1\rangle_S\langle 1| + p|2\rangle_S\langle 2|, \quad (15)$$

where we wish to shift a small amount δp of probability from $|0\rangle_S$ to $|2\rangle_S$. To do this, we take a thermal qubit with excitation probability $(p + \delta p)/(1 - q)$, so that the *ratio* of occupation probabilities in the bath qubit is the same as in that in the desired final state of the system. We then apply the unitary

$$|0\rangle_B|2\rangle_S|x\rangle_w \leftrightarrow |1\rangle_B|0\rangle_S|x + E_2 - E_B\rangle_w \quad (16)$$

where E_2 is the energy of the state $|2\rangle_S$, and all orthogonal states are left unchanged. We show in Appendix E that the change in the average energy of the weight during this process is equal to the free energy change of the system, to first order in δp .

From this it is straightforward to extend the qubit protocol to arbitrary diagonal states. For sufficiently large N we can choose a sequence of $N + 1$ states for the system where two adjacent states differ in occupation probability between a pair of levels by $\mathcal{O}(1/N)$, with the first and last states equal to the initial state of the system and its thermal state respectively⁷. We then pick the Hamiltonian for N thermal qubits such that unitaries of the form (16) take the reduced state of the system from one state in the sequence to the next. This is done by taking the thermal qubits to have populations equal to those of the pair of levels of interest of the next state in the sequence, after renormalising. In the limit $N \rightarrow \infty$ the average amount of work put into the weight equals the free energy change of the system, regardless of the precise choice of path.

Interestingly, an equivalent protocol would work for transitions between any two diagonal states, ρ and σ , extracting an average amount of work equal to $F_\beta(\rho) - F_\beta(\sigma)$. This implies that creating a diagonal state ρ from a thermal state costs the same average amount of work as can be extracted from ρ when it is thermalised. In this sense thermodynamic transitions between diagonal states are reversible⁸. This differs from the results of [16], who show that such transitions are irreversible when considering (almost) deterministic work extraction, rather than average work.

⁷ For example, we could first shift probability from all energy levels with higher probability in ρ_S than in $\tau_\beta(H_S)$ to the $|0\rangle_S$ state, then move probability from $|0\rangle_S$ to the remaining levels, using $N/(d - 1)$ steps for each pair of levels.

⁸ However, note that if a state is thermalised and then recreated using our protocol, the fluctuations in the position of the weight will increase.

E. Proof of optimality

We now show that the above protocol is optimal. In particular, we will argue that the maximum free energy that can be extracted from a transition between diagonal states ρ and σ is $F_\beta(\rho) - F_\beta(\sigma)$, using a modification of our proof of the second law (in Section III).

Suppose that we have a protocol which extracts average work $F_\beta(\rho) - F_\beta(\sigma) + \delta$ (where $\delta > 0$) when the system is transformed from ρ and σ . We can then use the above protocol to return the state from σ to ρ in a finite number of steps, extracting work $F_\beta(\sigma) - F_\beta(\rho) - \epsilon$, where we choose the number of thermal qubits such that ϵ is in the range $0 < \epsilon < \delta/2$. The net effect is that a positive average work $\geq \delta/2$ is extracted into the weight, and the system begins and ends the combined procedure in the same state ρ .

Considering total energy conservation for the combined protocol, we note that the average energy of the bath must have decreased by at least $\delta/2$, which means that the bath's entropy must also have decreased by a finite amount. However, the system does not change in entropy, and the change in entropy of the weight can be made as small as desired by choosing its initial state to be a sufficiently broad pure state (see Appendix A). Hence entropy conservation cannot be satisfied by our hypothetical protocol, and thus no protocol exists which extracts more average work than the free energy change.

V. EXTRACTING WORK FROM COHERENCE

So far all of our discussions have involved only states which are diagonal in the energy eigenbasis and the protocol presented involved shifting probability only between these energy eigenstates. We can of course consider more general states, and ask the question of how to extract the maximum amount of work from such states.

A. Motivating example

As a motivating example, let us consider a system with Hamiltonian $H_S = E_S|1\rangle_S\langle 1|$ prepared initially in the pure state

$$|\psi\rangle_S = \sqrt{1 - p_{eq}}|0\rangle_S + \sqrt{p_{eq}}|1\rangle_S \quad (17)$$

This state has the same average energy as the thermal state $\tau_\beta(H_S)$, however, since this state is pure, it has zero entropy and hence its free energy differs from that of the thermal state. Thus one would like to extract work from this state, with the limit on extractable work given by

$$W \leq F(\psi) - F(\tau_\beta(H_S)) = TS_c(p_{eq}) \quad (18)$$

Previously we were able to extract work from single copies of diagonal states. However, we show in the next section

that it is impossible to extract work from a single copy of $|\psi\rangle_s$ given the assumptions of the framework. This is similar to the analysis of [16] where they also found that the state (17) is useless for (almost) deterministic work extraction.

However, it is only a single copy which is a useless resource; two copies can be used to extract work, although the amount is less than that given by (18). The important point is the way in which work is extracted, which will suggest a different way in which one can understand an asymptotic protocol (in the number of system qubits) able to extract the optimal amount of work.

Consider two copies of the state $|\psi\rangle_s$, which written out in full is given by

$$|\psi\rangle_s^{\otimes 2} = (1 - p_{eq})|0\rangle_s|0\rangle_s + \sqrt{p_{eq}(1 - p_{eq})}(|0\rangle_s|1\rangle_s + |1\rangle_s|0\rangle_s) + p_{eq}|1\rangle_s|1\rangle_s \quad (19)$$

Let us consider that we now dephase this state in the energy eigenbasis, so called collective-dephasing. We can achieve this in two ways, either by allowing the state to evolve under its Hamiltonian freely for some unknown duration of time, or more systematically by applying controlled-Z operations between maximally mixed (therefore degenerate) bath qubits and the system. Note that both of these are energy conserving unitaries. In either case, at the end of the dephasing the state ρ is

$$\rho = (1 - p_{eq})^2|0\rangle_s|0\rangle_{ss}\langle 0|_s\langle 0| + 2p_{eq}(1 - p_{eq})|\psi^+\rangle_{ss}\langle\psi^+| + p_{eq}^2|1\rangle_s|1\rangle_{ss}\langle 1|_s\langle 1| \quad (20)$$

where $|\psi^+\rangle_s = \frac{1}{\sqrt{2}}(|0\rangle_s|1\rangle_s + |1\rangle_s|0\rangle_s)$. We note that since $|0\rangle_s|1\rangle_s$ and $|1\rangle_s|0\rangle_s$ have the same energy, the definite phase between them is preserved. The important point is that when considering the combined system we find that we have a two-dimensional degenerate subspace and within this subspace we are in the pure state $|\psi^+\rangle_s$, and *not* the orthogonal state $|\psi^-\rangle_s \equiv \frac{1}{\sqrt{2}}(|0\rangle_s|1\rangle_s - |1\rangle_s|0\rangle_s)$. Thus we have knowledge that we are with certainty in one of two orthogonal states and we can use this information to extract work from a thermal bath by allowing it to ‘isothermally expand’ into the full subspace. Randomising these two states brings the pair of qubits to their thermal state. We can thus apply the analogue of the protocol in Sec. IV C, coupling only the two states inside the degenerate subspace to the thermal bath, and it follows immediately that the final average energy of the weight, in the asymptotic limit, is

$$\langle E \rangle_w = 2p_{eq}(1 - p_{eq})T \log 2 \quad (21)$$

Except for the limiting case $p_{eq} = 0$, corresponding to the unphysical condition $E_s \rightarrow \infty$, this value is strictly less than the change in free energy of two system qubits,

$$2p_{eq}(1 - p_{eq})T \log 2 < 2TS_c(p_{eq}) \quad (22)$$

This inequality is due to the fact that the collective-dephasing changed the entropy of the system but extracted no work in the process. One may wonder therefore why this operation is considered in the first place. In Appendix F we show that we can apply the same analysis, including the collective-dephasing, to n copies of $|\psi\rangle_s$ and extract, in the asymptotic limit $n \rightarrow \infty$, work equal to the free energy change $nTS_c(p_{eq})$. This is possible since, as we also show, collective-dephasing increases the entropy of n systems by an amount which grows as $\log(n + 1)$, which becomes negligible in the asymptotic limit. We thus see that coherence between states of different energy in fact plays no thermodynamic role asymptotically in work extraction⁹.

Altogether this highlights an essential difference between diagonal and non-diagonal states; When $E_s \neq 0$, the state of multiple qubits in a diagonal state will be maximally mixed inside each energy eigenspace. However, for non-diagonal states, these subspaces take on non-trivial states, and work can be extracted from them by allowing them to ‘isothermally expand’ into the full subspace.

B. General work extraction protocol

We now present a general strategy to extract work from n copies of an arbitrary state ρ , of arbitrary dimension d . It is helpful to divide the protocol into three steps:

- (i) Collectively dephase the state $\rho^{\otimes n}$ in the joint energy eigenbasis.
- (ii) Allow the resulting state to ‘isothermally expand’ inside each energy eigenspace, until it equals $\omega^{\otimes n}$, where ω is the state obtained by dephasing ρ in its energy eigenbasis.
- (iii) Individually transform each copy of ω into $\tau_\beta(H_S)$.

It follows from the results of Sec. IV D that steps (2) and (3) can be performed whilst extracting an average amount of work equal to the free energy change. The free energy change in the first step is not captured as work, but is simply lost. However, we show in the Appendix F that the amount of work lost in this step is at most $T(d - 1)\log(n + 1)$.

The average amount of work extracted per copy of the original system is therefore

$$\delta W \geq F_\beta(\rho) - F_\beta(\tau_\beta(H_S)) - \frac{T(d - 1)}{n} \log(n + 1). \quad (23)$$

In the limit as $n \rightarrow \infty$, we therefore extract an average work per copy of ρ equal to its free energy change, as desired.

If we only have a single copy of the system, the amount of work we can extract from ρ is given by the free energy

⁹ In the other direction, that of creating non-thermal states from a supply of work, in [23] it was shown that a phase reference (e.g. a large coherent state) is consumed at a sub-linear rate in this process.

change of the decohered state, $F_\beta(\omega) - F_\beta(\tau_\beta(H_S))$. This can be obtained by first dephasing ρ into ω and then using the protocol in Sec. IV D. We now show that this is optimal, following the approach given in [16].

Firstly, note that as we require the average work extracted by a protocol to be independent of the initial state of the weight, we are free to choose that state however we like – here we choose it to be a very narrow wavepacket centred on zero. Now consider a decomposition of the total state space into subspaces, each of which has total energy (of the system, bath and weight) close to one of the energy eigenvalues of the system and bath¹⁰. Note that any work extraction protocol can be followed by a transformation which decoheres the state with respect to these total energy subspaces, without affecting the average work extracted. However, this decohering operation commutes with the unitaries used in the protocol, so we can move it to the beginning of the protocol without changing the work extracted. At the beginning of the protocol, this operation has the sole effect of decohering the system in its energy eigenbasis (changing ρ to ω). Hence the protocol extracts the same amount of work as it would have if it had operated on ω .

VI. PERSPECTIVE USING ‘VIRTUAL TEMPERATURES’

Consider the model of the smallest possible heat engines [4, 33], consisting of two qubits, one with energy spacing E_1 in thermal contact with a bath at temperature T_1 , and a second with energy spacing E_2 in contact with a bath at T_2 . As a composite system this pair of qubits is not in a global thermal state. Nevertheless, let us assign a ‘virtual temperature’ to each transition via the identification [33]

$$\frac{p(e)}{p(g)} = e^{-\beta_v(E_e - E_g)} \quad (24)$$

where $p(e)$ and $p(g)$ are the probability of being in the ‘excited’ state (the state of higher energy) and ground state (the state of lower energy) respectively, with E_e and E_g being their energy eigenvalues, and $\beta_v = 1/T_v$ is the ‘inverse virtual temperature’. For the case in hand, the virtual inverse temperature for the states $|0\rangle_1|1\rangle_2$ and $|1\rangle_1|0\rangle_2$ is

$$\beta_v = \frac{S'_c(p_1) - S'_c(p_2)}{E_1 - E_2} = \frac{E_1\beta_1 - E_2\beta_2}{E_1 - E_2} \quad (25)$$

where we have denoted the probability for qubit i to be in the excited state by p_i , and we have used the relation $S'_c(p) = \beta E$ for a thermal state.

The motivation for introducing the concept of virtual temperatures is that for the smallest machines it gives a concise characterisation of their behaviour; the function of the

machine is determined by the virtual temperature – refrigerator, heat pump or heat engine, if colder, hotter, or negative temperature respectively; Carnot efficiency is approached when the virtual temperature approaches the external temperature for the refrigerator and pump, and when approaching minus infinite temperature for the heat engine.

It is the final machine mentioned, the heat engine, which is relevant here, since this is the machine which produces work. We can think of the system qubit and bath qubit together to be a two-qubit heat engine and that the operating point of the protocol is the Carnot point of the engine.

Let us return to the motivating example of Section IV A. Our system qubit had spacing E_S with excited state probability p , whilst the bath qubit had spacing E_B with excited state probability q . Let us rename the state $|1\rangle_B|0\rangle_S = |0\rangle_v$ and $|0\rangle_B|1\rangle_S = |1\rangle_v$. With this renaming, the interaction (7) becomes

$$|1\rangle_v|x\rangle_w \leftrightarrow |0\rangle_v|x + E_v\rangle_w \quad (26)$$

where $E_v = E_S - E_B$ is the spacing of the virtual qubit. Thus from this perspective it is explicit that the unitary is the one which exchanges energy from the virtual qubit into the weight, analogous to the Hamiltonian of the smallest heat engine [33]. The operating point of the protocol was to choose E_B such that $q = p - \delta p$. In this regime, to first order, the inverse virtual temperature becomes

$$\beta_v \simeq \delta p \frac{S''_c(p)}{E_S - T S'_c(p)} \quad (27)$$

where $S''_c(p) = -\frac{1}{1-p} - \frac{1}{p} \leq 0$. Thus the inverse virtual temperature becomes vanishingly small and negative as $\delta p \rightarrow 0$, which is exactly the Carnot limit of the small heat engine.

The picture that one obtains from this perspective is that the reason why any state which is not a thermal state at temperature T is a resource is because the *composite system of resource state + thermal state* contain non-trivial virtual temperatures. One can extract work from a system approaching its free energy when the only exchanges between the composite system and weight involve virtual qubits whose inverse temperatures become negatively infinite.

Finally, we conclude by noting the protocol for systems with many energy levels becomes easier to explain using this perspective. A non-thermal state will have transitions (i.e. pairs of levels) at many different temperatures, and each pair needs to be brought to the same temperature in the thermalisation process. In the protocol the bath qubits interact only with a single pair of levels at a time. To approach reversibility it must be the case that any systems which interact only differ in temperature by an infinitesimal amount, but this is exactly the condition enforced by choosing the renormalised probability of the pair to be close to the bath qubit populations. Finally, the unitary which is applied is the basic thermalising unitary and the energy exchanged in the limit is precisely the free energy.

¹⁰ In particular the i^{th} subspace corresponds to the total energy E lying in the range $\frac{E_i - E_{i-1}}{2} \leq E \leq \frac{E_{i+1} - E_i}{2}$, where E_i are the energy eigenvalues of the system and bath. Furthermore we choose the width of the weight’s initial wavepacket to be narrower than the smallest subspace.

VII. CONCLUSIONS

To summarise, in this paper we have analysed the fundamental laws of thermodynamics using the recently proposed concept of resource theories. We proposed an alternative resource theory which differs in certain aspects from the original one. In this conceptual framework we have proved a number of results. First we proved the second law of thermodynamics. Second we proved that for systems that are diagonal in the energy eigenbasis, but not at equilibrium with the thermal bath, we can extract work equal to the free energy by processing each system *individually*, hence showing the free energy is a relevant quantity for individual systems. Third, we proved that when systems are not diagonal in the energy eigenbasis, i.e. they have coherences between different energy eigenstates, it is not possible to extract work equal to the free energy by processing them individually. Work equal to the free energy can however be extracted if we allow for collective processing of such systems (in the asymptotic limit of many copies).

A number of important questions open immediately. First of all, whether the assumptions which underlie our resource theory framework are too restrictive. In particular, this may be the reason why we cannot extract the entire free energy by processing single systems with coherences between different energy levels. Indeed, given our assumptions we are very limited in how we can access and manipulate the relative phase between two different energy levels, hence we cannot interfere them in a controlled way. As such, for all practical purposes the different energy levels behave as if they were decohered; it is precisely the free energy of this decohered state that is the maximum we can extract in our framework. To be able to interfere different energy levels, i.e. to control their relative phase, we need a “frame of reference” for it. One way to look at this is that we need some other resource with coherent superpositions between energy levels¹¹. Another way to express this is that we essentially need a clock: the phase between two energy levels varies in time, so a “frame of reference” for it is essentially a clock. When collectively processing many copies the different systems act effectively as reference systems for each other. For individual processing however we essentially need to add an

external clock. In our present framework we did not include such a clock since we were not able to guarantee that one cannot unwittingly cheat by using the clock itself as a source of free energy. It is conceivable however that a more careful formulation of the resource theory would permit integrating a clock in a consistent and controlled manner. In this case, it may be possible that work equal to the free energy can be extracted from all quantum states by only processing individual copies.

Second, while in the above we asked if we can relax some of the constraints imposed in the resource theory, for example by including a clock, we must also enquire whether we may indeed need to strengthen some of the constraints. In particular, one may be concerned that we have used unitary transformations as opposed to Hamiltonian evolution. Unitaries imply controlled evolution, i.e. turning on and off interactions at precise time intervals. One may wonder whether this does not require some free energy to achieve. If this is the case, then it is only an illusion that we have extracted work equal to the free energy from the system, since we had to pay to implement the process. Although such a detailed analysis is missing, the fact that all of the results obtained within the resource theory framework so far are consistent with our expectations is an implicit indication that the assumptions are indeed justified.

To conclude, the resource theory framework seems to be a natural and powerful way to approach thermodynamics. It has already delivered significant results and we believe that further investigation along these lines will lead to a much deeper understanding of the foundations of thermodynamics.

ACKNOWLEDGMENTS

SP and PS acknowledge support from the European project (Integrated Project “Q-ESSENCE”). SP acknowledges support from the ERC (Advanced Grant “NLST”), and the Templeton Foundation. AJS acknowledges support from the Royal Society. PS is grateful to Lidia del Rio, Johan Åberg, Joe Renes, Philippe Faist, Renato Renner, Oscar Dahlsten and Luis Masanes for interesting discussions.

-
- [1] J. D. Bekenstein, *Black Holes and Entropy*, Phys. Rev. D **7**, 2333 (1973).
 - [2] J. M. Bardeen, B. Carter, and S. W. Hawking, *The four laws of black hole mechanics*, Comm. Math. Phys. **31**, 161 (1973).
 - [3] E. Geva and R. Kosloff, *The quantum heat engine and heat pump: An irreversible thermodynamic analysis of the three-level amplifier*, J. Chem. Phys. **104**, 7681 (1996).
 - [4] M. Youssef, G. Mahler, and A.-S. F. Obada, *Quantum optical*

thermodynamic machines: Lasing as relaxation, Phys. Rev. E **80**, 061129 (2009).

- [5] F. Tonner and G. Mahler, *Autonomous quantum thermodynamic machines*, Phys. Rev. E **72**, 066118 (2005).
- [6] N. Linden, S. Popescu, and P. Skrzypczyk, *How Small Can Thermal Machines Be? The Smallest Possible Refrigerator*, Phys. Rev. Lett. **105**, 130401 (2010).
- [7] G. Gemma, M. Michel, and G. Mahler, *Quantum Thermodynamics*, Springer, New York, 2004.
- [8] E. P. Gyftopoulos and G. P. Beretta, *Thermodynamics: Foundations and Applications*, Dover, New York, 2005.
- [9] I. Procaccia and R. D. Levine, *Potential work: A statistical-mechanical approach for systems in disequilibrium*, J.

¹¹ Note that the weight cannot be used for this purpose because of our assumption that the work extracted must be independent of its initial state

- Chem. Phys. **65**, 3357 (1976).
- [10] R. Alicki, M. Horodecki, P. Horodecki, and R. Horodecki, *Thermodynamics of Quantum Information Systems – Hamiltonian Description*, Open Syst. Inf. Dyn. **11**, 205 (2004).
 - [11] H.-H. Hasegawa, J. Ishikawa, and D. J. Driebe, *Generalization of the second law for a nonequilibrium initial state*, Phys. Lett. A **374**, 1001 (2010).
 - [12] K. Takara, H.-H. Hasegawa, and D. J. Driebe, *Generalization of the second law for a transition between nonequilibrium states*, Phys. Lett. A **375**, 88 (2010).
 - [13] M. Esposito and C. Van den Broeck, *Second law and Landauer principle far from equilibrium*, Europhys. Lett. **95**, 40004 (2011).
 - [14] J. Åberg, *Truly work-like work extraction*, arXiv:1110.6121 [quant-ph] (2011).
 - [15] G. Verley and D. Lacoste, *Fluctuation theorems and inequalities generalizing the second law of thermodynamics out of equilibrium*, Phys. Rev. E **86**, 051127 (2012).
 - [16] M. Horodecki and J. Oppenheim, *Fundamental limitations for quantum and nano thermodynamics*, arXiv:1111.3834 [quant-ph] (2011).
 - [17] C. H. Bennett, H. Bernstein, S. Popescu, and B. Schumacher, *Concentrating partial entanglement by local operations*, Phys. Rev. A **53**, 2046 (1996).
 - [18] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, *Quantum entanglement*, Rev. Mod. Phys. **81**, 865 (2009).
 - [19] S. D. Bartlett, T. Rudolph, and R. W. Spekkens, *Reference frames, superselection rules, and quantum information*, Rev. Mod. Phys. **79**, 555 (2007).
 - [20] M. Horodecki, P. Horodecki, and J. Oppenheim, *Reversible transformations from pure to mixed states and the unique measure of information*, Phys. Rev. A **67**, 062104 (2003).
 - [21] I. Marvian and R. W. Spekkens, *The theory of manipulations of pure state asymmetry I: basic tools and equivalence classes of states under symmetric operations*, arXiv:1104.0018 [quant-ph] (2011).
 - [22] M. Horodecki and J. Oppenheim, *(Quantumness in the context of) Resource Theories*, arXiv:1209.2162 [quant-ph] (2012).
 - [23] F. G. S. L. Brandao, M. Horodecki, J. Oppenheim, J. M. Renes, and R. W. Spekkens, *The Resource Theory of Quantum States Out of Thermal Equilibrium*, arXiv:1111.3882 [quant-ph] (2011).
 - [24] D. Janzing, P. Wocjan, R. Zeier, R. Geiss, and Th. Beth, *Thermodynamic Cost of Reliability and Low Temperatures: Tightening Landauer's Principle and the Second Law*, Int. J. Th. Phys. **39**, 2717 (2000).
 - [25] O. C. O. Dahlsten, R. Renner, E. Rieper, and V. Vedral, *Inadequacy of von Neumann entropy for characterizing extractable work*, New J. Phys. **13**, 053015 (2011).
 - [26] L. d. Rio, J. Åberg, R. Renner, O. C. O. Dahlsten, and V. Vedral, *The thermodynamic meaning of negative entropy*, Nature **474**, 61 (2011).
 - [27] D. Egloff, O. C. O. Dahlsten, R. Renner, and V. Vedral, *Laws of thermodynamics beyond the von Neumann regime*, arXiv:1207.0434 [quant-ph] (2012).
 - [28] P. Faist, F. Dupuis, J. Oppenheim, and R. Renner, *A Quantitative Landauer's Principle*, arXiv:1211.1037 [quant-ph] (2012).
 - [29] E. H. Lieb and J. Yngvason, *The physics and mathematics of the second law of thermodynamics*, Phys. Rept. **310**, 1 (1999).
 - [30] J. Ladyman, S. Presnell, and A. J. Short, *The use of the information-theoretic entropy in thermodynamics*, Stud. Hist. Phil. Sci. B **39**, 315 (2008).
 - [31] L. Szilard, *über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen*, Z. Phys. **53**, 840 (1929).
 - [32] H. S. Leff and A. F. Rex (eds.), *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing*, 2nd ed., Institute of Physics, Bristol, 2003.
 - [33] N. Brunner, N. Linden, S. Popescu, and P. Skrzypczyk, *Virtual qubits, virtual temperatures, and the foundations of thermodynamics*, Phys. Rev. E **85**, 051111 (2012).
 - [34] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge, UK, 2000.

APPENDICES

Appendix A: Independence of the work on the initial state of the weight

In each step of our protocols in which work is extracted, the state of a thermal qubit is swapped with a pair of energy levels in the system, with a corresponding translation of the weight in order to conserve total energy. Consequently, each work-extraction step can be represented by a unitary transformation of the form

$$U = \sum_{ij} \Pi_{ij} |i\rangle\langle j| \otimes \Gamma_{E_j - E_i} \quad (\text{A1})$$

where the states $\{|i\rangle\}$ are an orthonormal basis of energy eigenstates for the system and the relevant portion of the bath, such that $(H_B + H_S)|i\rangle = E_i|i\rangle$, Γ_a is the translation operator on the weight, and Π_{ij} is a permutation matrix. Furthermore, apart from the initial decohering of non-classical states (discussed in Section V), which is entirely independent from the weight, the entire protocol can be cast in this form.

It is easy to see that all such unitaries commute with translations on the weight. We now show that the average work extracted by a unitary of this form does not depend on the initial state of the weight (even if it is initially correlated with the state of the system). Let us denote the initial state of the system, bath and weight by the density operator ρ . The average work extracted is given by

$$\begin{aligned} \langle W \rangle &= \text{tr}(H_W U \rho U^\dagger) - \text{tr}(H_W \rho), \\ &= \text{tr}((U^\dagger H_W U - H_W) \rho) \end{aligned} \quad (\text{A2})$$

where $H_W = \mathbb{1}_{SB} \otimes \hat{x}_w$ is the Hamiltonian of the weight (defined for convenience as an operator on the entire system).

Now note that

$$\begin{aligned}
U^\dagger H_W U &= \sum_{ijkl} \Pi_{kl} \Pi_{ij} |l\rangle\langle k| |i\rangle\langle j| \otimes \left(\Gamma_{E_k - E_l} \hat{x} \Gamma_{E_j - E_i} \right) \\
&= \sum_{ijl} \Pi_{il} \Pi_{ij} |l\rangle\langle j| \otimes \left(\Gamma_{E_i - E_l} \hat{x} \Gamma_{E_j - E_i} \right) \\
&= \sum_{ij} \Pi_{ij} |j\rangle\langle j| \otimes \left(\Gamma_{E_i - E_j} \hat{x} \Gamma_{E_j - E_i} \right) \\
&= \sum_{ij} \Pi_{ij} |j\rangle\langle j| \otimes \left(\hat{x} + (E_j - E_i) \mathbb{1}_w \right) \\
&= H_W + \sum_{ij} \Pi_{ij} (E_j - E_i) |j\rangle\langle j| \otimes \mathbb{1}_w \quad (\text{A3})
\end{aligned}$$

where in the third line we have used the fact that for permutation matrices $\Pi_{il} \Pi_{ij} = \Pi_{ij} \delta_{jl}$ (where δ_{jl} is the Kronecker delta function), and in the fourth line we have used the fact that $\Gamma_{-a} \hat{x} \Gamma_a = \hat{x} + a \mathbb{1}_w$. Inserting this expression in (A2), we obtain

$$\langle W \rangle = \text{tr}_{SB} \left(\left(\sum_{ij} \Pi_{ij} (E_j - E_i) |j\rangle\langle j| \right) \rho_{SB} \right) \quad (\text{A4})$$

where $\rho_{SB} = \text{tr}_w(\rho)$ is the reduced density matrix of the system and bath. Hence the average amount of work extracted is independent of the initial state of the weight as desired.

Note that the most general energy-conserving unitary transformations which commute with all translations on the weight have the form

$$U = \sum_{ij} u_{ij} |i\rangle\langle j| \otimes \Gamma_{E_j - E_i} \quad (\text{A5})$$

where u_{ij} is an arbitrary unitary matrix. However, for such unitaries the work extracted would generally depend on the initial state of the weight.

Appendix B: Proof of second law of thermodynamics

In this appendix, we provide further details for our proof of the second law in Section B.

We first argue that any reduction in the average energy of an initially thermal state (with positive temperature) must also yield a reduction in its entropy. We will only need to use the fact that a thermal state is the maximal entropy state with a given average energy. If a system starts in a thermal state and is transformed to a final state with fixed average energy, the entropy change ΔS_B will be maximised when the final state is also thermal. In the case where the average energy decreases, and the initial state has positive temperature, the final state will be a thermal state of lower temperature, and thus lower entropy.

We now show that the entropy change of the weight in any particular protocol may be made as small as desired by choosing its initial state to be a sufficiently broad pure state (e.g. a state $|\psi_L\rangle$ with wavefunction $\psi_L(x)$ equal to

$1/\sqrt{2L}$ for $|x| \leq L$, and 0 otherwise, where $L \gg 1$). First note that the requirement that transformations are translationally invariant on the weight means that they have the form given in (A5). When applied to a thermal state of the bath, the effect of the protocol on the weight is to transform it into a finite mixture of translations of the initial state $\rho_w = \sum_{ij} p_{ij} \Gamma_{E_i - E_j} |\psi_L\rangle\langle\psi_L| \Gamma_{E_j - E_i}$, where p_{ij} are probabilities and i and j label energy eigenstates (of energies E_i and E_j) of the bath. If the thermal systems used in the protocol have total dimension d , and the initial state of the weight is pure, then the final state of the weight lies in a d^2 -dimensional subspace spanned by the states $\Gamma_{E_i - E_j} |\psi_L\rangle\langle\psi_L|$ for $i, j \in \{1, 2, \dots, d\}$.

As the final state of the weight lives in a finite dimensional subspace, its entropy can be shown to be continuous due to Fannes' inequality [34].

$$|S(\rho_w) - S(|\psi_L\rangle\langle\psi_L|)| \leq D \log \left(\frac{d^2}{D} \right) \quad (\text{B1})$$

where D is the trace distance between ρ_w and $|\psi_L\rangle\langle\psi_L|$. We can make D as small as we like by taking sufficiently large L and therefore make the entropy change as small as we like. For the case in hand, a straightforward calculation shows that $D \leq \sqrt{\frac{a}{L}}$, where $a = \max_{ij} |E_i - E_j|$ is the maximum energy gap of the system-bath Hamiltonian.

Note that all of the above arguments apply equally well if we consider a system in addition to the bath which is initially in a mixture of energy eigenstates, as considered in Section IV E.

Appendix C: Motivating example details

In this appendix we shall provide more details into the example of Section IV A, where a state undergoes a transformation and in the process supplies an amount of work W .

Consider the state

$$\rho_S = (1-p)|0\rangle_S\langle 0| + p|1\rangle_S\langle 1| \quad (\text{C1})$$

with Hamiltonian $H_S = E_S|1\rangle_S\langle 1|$, along with a thermal qubit

$$\tau_B(H_B) = (1-r)|0\rangle_B\langle 0| + r|1\rangle_B\langle 1| \quad (\text{C2})$$

with Hamiltonian $H_B = E_B|1\rangle_B\langle 1|$, where $E_B = TS'_c(r)$. Let us bring in a weight in the state $|0\rangle_w$, and apply the unitary transformation U , which interchanges

$$|0\rangle_B |1\rangle_S |x\rangle_w \leftrightarrow |1\rangle_B |0\rangle_S |x + E_S - E_B\rangle_w \quad (\text{C3})$$

whilst leaving all other orthogonal states unchanged. Apply-

ing this unitary to the combined system we find that

$$\begin{aligned}\sigma_{BSw} &= U \tau_\beta(H_B) \otimes \rho_S \otimes |0\rangle_w \langle 0| U^\dagger \\ &= r p |1\rangle_B |1\rangle_S |0\rangle_w \langle 1|_S \langle 1|_w \langle 0| \\ &\quad + r(1-p) |0\rangle_B |1\rangle_S |E_B - E_S\rangle_w \langle 0|_S \langle 1|_w \langle E_B - E_S| \\ &\quad + (1-r)p |1\rangle_B |0\rangle_S |E_S - E_B\rangle_w \langle 1|_S \langle 0|_w \langle E_S - E_B| \\ &\quad + (1-r)(1-p) |0\rangle_B |0\rangle_S |0\rangle_w \langle 0|_S \langle 0|_w \langle 0| \quad (C4)\end{aligned}$$

From this one can find that the final reduced state of the system and bath qubit are $\rho' = \text{tr}_{Bw} \sigma = \tau_\beta(H_B)$ and $\tau' = \text{tr}_{Sw} \sigma = \rho_S$ respectively. Thus we see that at the level of reduced states these two systems have swapped states. The change in free energy of the system ΔF_β is

$$\begin{aligned}\Delta F_\beta &= F_\beta(\rho) - F_\beta(\tau_\beta(H_B)) \\ &= (pE_S - TS_c(p)) - (rE_S - TS_c(r)) \\ &= (p-r)E_S - T(S_c(p) - S_c(r)) \quad (C5)\end{aligned}$$

Moving on to the weight its final reduced state Ω'_w is

$$\begin{aligned}\Omega'_w &= (rp + (1-r)(1-p)) |0\rangle_w \langle 0| \\ &\quad + r(1-p) |E_B - E_S\rangle_w \langle E_B - E_S| \\ &\quad + (1-r)p |E_S - E_B\rangle_w \langle E_S - E_B| \quad (C6)\end{aligned}$$

The change in average energy of the weight $\Delta \langle E \rangle_w$ is thus

$$\begin{aligned}\Delta \langle E \rangle_w &= \text{tr}(H_w \Omega'_w) - \text{tr}(H_w |0\rangle_w \langle 0|) \\ &= (p-r)(E_S - E_B) \\ &= (p-r)(E_S - TS_c(r)) \quad (C7)\end{aligned}$$

Now, let us take $r = p - \delta p$ and expand to second order in δp both ΔF_β and $\Delta \langle E \rangle_w$,

$$\Delta F_\beta = \delta p(E_S - TS'_c(p)) + \delta p^2 \frac{TS''_c(p)}{2} + \mathcal{O}(\delta p^3) \quad (C8)$$

$$\Delta \langle E \rangle_w = \delta p(E_S - TS'_c(p)) + \delta p^2 TS''_c(p) + \mathcal{O}(\delta p^3) \quad (C9)$$

The change in free energy and the change in average energy of the weight thus coincide to first order in δp and differ only by a factor of 2 in the second order term. Since $S''_c(p) < 0$, we see immediately that $\Delta \langle E \rangle_w \leq \Delta F_\beta$, so that the change in free energy is always larger than the change in average energy of the weight, as we would expect.

Appendix D: Calculating the average energy and spread

In this appendix we will show how to arrive at equation (11) for the asymptotic state of the ladder. To do so let us consider first those instances where the system qubits starts in state $|0\rangle_S$. Given this we shall then calculate the expected energy of the weight and its variance in the limit of large N to first order, from which we shall be able to conclude that it approaches a definitive value in the asymptotic limit at the

desired energy. A analogous method can also be employed assuming the system starts instead in the state $|1\rangle_S$, with the desired result also obtained.

By calculating the first few stages of the protocol and then by induction it is easy to show that the expected energy of the weight at the end of the protocol is

$$\langle E \rangle_w^{[0]} = -p_{eq} E_S + p E_B^{(1)} + \sum_{k=1}^N (q^{(k)} - q^{(k-1)}) E_B^{(k)} \quad (D1)$$

where the superscript $[0]$ is a reminder that this is the case for the system starting in the state $|0\rangle_S$, and as defined in the main text we have

$$q^{(k)} = p + \frac{k}{N}(p_{eq} - p) \quad E_B^{(k)} = \frac{1}{\beta} \log \left(\frac{1-q^{(k)}}{q^{(k)}} \right) \quad (D2)$$

By expanding the logarithms in powers of $\frac{k}{N}$ and denoting $a = (p_{eq} - p)$, after some basic manipulations this can be re-written as

$$\begin{aligned}\langle E \rangle_w^{[0]} &= -p_{eq} E_S + p E_B^{(1)} + \frac{a}{\beta} \log \left(\frac{1-p}{p} \right) \\ &\quad + \frac{a}{\beta N} \sum_{k=1}^N \sum_{n=1}^{\infty} \left(-\frac{1}{n} \left(\frac{ak}{(1-p)N} \right)^n + \frac{1}{n} \left(\frac{-ak}{pN} \right)^n \right) \quad (D3)\end{aligned}$$

The advantage of performing this expansion is that we can now interchange the order of summation and note that the only term which depends upon k is the term k^n which is common to both terms in the sum. We can thus use the expansion

$$\sum_{k=1}^N k^n = \frac{N^{n+1}}{n+1} + \frac{N^n}{2} + \mathcal{O}(N^{n-1}) \quad (D4)$$

to write

$$\begin{aligned}\langle E \rangle_w^{[0]} &= -p_{eq} E_S + p E_B^{(1)} + \frac{a}{\beta} \log \left(\frac{1-p}{p} \right) \\ &\quad + \frac{1}{\beta} \sum_{n=1}^{\infty} \left(\frac{a^{n+1}}{n} \left(\frac{1}{(-p)^n} - \frac{1}{(1-p)^n} \right) \left(\frac{1}{n+1} + \frac{1}{2N} \right) \right) \\ &\quad + \mathcal{O} \left(\frac{1}{N^2} \right) \quad (D5)\end{aligned}$$

This sum can now be evaluated explicitly, and after re-expressing E_S in terms of $S'_c(p_{eq})$, expanding $E_B^{(1)}$ to first order in powers of $\frac{1}{N}$, and an amount of manipulation we arrive at

$$\begin{aligned}\langle E \rangle_w^{[0]} &= \frac{1}{\beta} \log \left(\frac{1-p}{1-p_{eq}} \right) + \frac{1}{\beta} (p_{eq} - p) \left(\frac{1}{2} (S'_c(p_{eq}) \right. \\ &\quad \left. - S'_c(p)) - \frac{1}{1-p} \right) \frac{1}{N} + \mathcal{O} \left(\frac{1}{N^2} \right) \quad (D6)\end{aligned}$$

thus demonstrating that the weight is peaked around the point expected when the system starts off in the state $|0\rangle_S$ with an offset which vanishes in the asymptotic limit.

Considering the case now when the system starts off in the state $|1\rangle_s$ and again calculating the first few stages of the protocol and then by induction it can be shown that

$$\langle E \rangle_w^{[1]} = \langle E \rangle_w^{[0]} + E_S - E_B^{(1)} \quad (\text{D7})$$

Thus by re-expressing E_S in terms of $S'_c(p_{eq})$ and expanding $E_B^{(1)}$ to first order it is straightforward given the previous results to show that

$$\langle E \rangle_w^{[1]} = \frac{1}{\beta} \log \left(\frac{p}{p_{eq}} \right) + \frac{1}{\beta} (p_{eq} - p) \left(\frac{1}{2} (S'_c(p_{eq}) - S'_c(p)) + \frac{1}{p} \right) + \mathcal{O} \left(\frac{1}{N^2} \right) \quad (\text{D8})$$

which again gives the desired result. Let us move on now to a calculation of the variance $\Delta E_w^{2[i]} = \langle E^2 \rangle_w^{[i]} - \langle E \rangle_w^{2[i]}$ of each of the two packets, where for simplicity we will neglect the finite width of the initial wavepacket. Again by induction it can be shown that the spread of both packets is identical (and hence we shall drop the superscript $[i]$) and equal to

$$\Delta E_w^2 = \sum_{k=1}^N q^{(k)} (1 - q^{(k)}) (E_B^{(k+1)} - E_B^{(k)}) \quad (\text{D9})$$

Let us consider the k^{th} term in this sum, multiplied by N . Defining $x = \frac{k}{N}$ and expanding this summand in terms of $\frac{1}{N}$ we find

$$\begin{aligned} & q^{(k)} (1 - q^{(k)}) (E_B^{(k+1)} - E_B^{(k)}) \\ &= \frac{T^2 (p_{eq} - p)^2}{(p + x(p_{eq} - p))(1 - p - x(p_{eq} - p))N^2} + \mathcal{O} \left(\frac{1}{N^3} \right) \end{aligned} \quad (\text{D10})$$

Thus for large N the leading order contribution to this sum becomes equal to

$$\Delta E_w^2 \simeq \frac{1}{N} \int_0^1 \frac{T^2 (p_{eq} - p)^2 dx}{(p + x(p_{eq} - p))(1 - p - x(p_{eq} - p))} \quad (\text{D11})$$

Evaluating this integral explicitly we find finally that

$$\Delta E_w^2 = \frac{T^2 (p_{eq} - p) (S'_c(p_{eq}) - S'_c(p))}{N} + \mathcal{O} \left(\frac{1}{N^2} \right) \quad (\text{D12})$$

Thus we see that each peak has a spread of the order of $1/\sqrt{N}$ which vanishes in the asymptotic limit $N \rightarrow \infty$. Therefore in this limit we can conclude that the distribution of the ladder approaches a two-peaked distribution, as given by equation (11).

Appendix E: Extracting work from diagonal quantum systems of arbitrary dimension

In this appendix we wish to show that the average work extracted when moving the occupation probability between energy levels is still equal to the free energy in a d -dimensional system. It will suffice to consider a qutrit, and

moving probability δp from the ground state $|0\rangle_s$ to the second excited state $|2\rangle_s$.

Let us take our system to be initially in the state

$$\rho_S = (1 - p - q)|0\rangle_s \langle 0| + q|1\rangle_s \langle 1| + p|2\rangle_s \langle 2|, \quad (\text{E1})$$

and thus to end in the state

$$\rho'_S = (1 - p - q - \delta p)|0\rangle_s \langle 0| + q|1\rangle_s \langle 1| + (p + \delta p)|2\rangle_s \langle 2|. \quad (\text{E2})$$

where the system Hamiltonian is

$$H_S = E_1|1\rangle_s \langle 1| + E_2|2\rangle_s \langle 2|. \quad (\text{E3})$$

To achieve this, we append a bath qubit with energy level spacing E_B and ground state population r and apply the unitary transformation which interchanges only the states

$$|1\rangle_B |0\rangle_s |x\rangle_w \leftrightarrow |0\rangle_B |2\rangle_s |x + E_B - E_2\rangle_w \quad (\text{E4})$$

This induces on the system and weight the transformation

$$\begin{aligned} \rho_S \otimes |0\rangle_w \langle 0| &\rightarrow (1 - p - q)(1 - r)|0\rangle_s \langle 0|_w |0\rangle_s \langle 0|_w \\ &+ q|1\rangle_s \langle 0|_w |1\rangle_s \langle 0|_w + p r |2\rangle_s \langle 0|_w |2\rangle_s \langle 0|_w \\ &+ p(1 - r)|0\rangle_s \langle 1|_w |E_2 - E_B\rangle_w \langle 0|_w \langle E_2 - E_B|_w \\ &+ (1 - p - q)r |2\rangle_s \langle 1|_w |E_B - E_2\rangle_w \langle 2|_w \langle E_B - E_2|_w \end{aligned} \quad (\text{E5})$$

where we now carry out the final trace over the bath qubit for brevity of presentation. Thus in order that the final reduced state of the system is ρ'_S we find that the ground state population of the bath must be chosen to be

$$r = \frac{p + \delta p}{1 - q} \quad (\text{E6})$$

The denominator is the total probability to be in either of the states $|0\rangle_s$ or $|1\rangle_s$, and hence this expression can be seen as a renormalised version of the qubit case, where the total probability to be in either state always sums to unity. We see that it is the ratio of probabilities which must be close to the ratio of populations of the thermal qubits, which is equivalent to saying that they must be close in temperature (see Sec. VI).

The change in energy of the weight is given by

$$\begin{aligned} \delta W &= (p - (1 - q)r)(E_2 - E_B) \\ &= -\delta p (E_2 - E_B) \end{aligned} \quad (\text{E7})$$

thus using the fact that $E_B = \frac{1}{\beta} S'_c(r)$ and expanding to first order in δp we find that the work extracted from the system is given by

$$\delta W \simeq -\delta p \left(E_2 - \frac{1}{\beta} S'_c \left(\frac{p}{1 - q} \right) \right) \quad (\text{E8})$$

which a straightforward calculation verifies is equal to the change in free energy $F(\rho_S) - F(\rho'_S)$ to first order in δp . If we do a sequence of N such transitions between the states ρ and σ , in each of which $\delta p = \mathcal{O}(1/N)$, then we will obtain $\delta W = F(\rho_S) - F(\rho_f) + \mathcal{O}(1/N)$. In the asymptotic limit in which the number of thermal qubits N goes to infinity, we will therefore extract average free energy equal to the free energy change.

Appendix F: Work extraction from coherence

In this appendix we will present the full details for how work can be extracted from coherence. We shall first consider the special case presented in the main text for n systems, and then move on to the general case.

1. Motivating example

Consider n copies of $|\psi\rangle_s$, where

$$|\psi\rangle_s = \sqrt{1-p_{eq}}|0\rangle_s + \sqrt{p_{eq}}|1\rangle_s \quad (F1)$$

There will be $n+1$ energy eigenspaces with energies kE_s , $k = 0, \dots, n$. The dimension of the eigenspace with energy kE_s will be $\binom{n}{k}$. After collective-dephasing in each eigenspace the system will be a pure superposition over all states, and the probability to be in each eigenspace will be $\binom{n}{k} p_{eq}^k (1-p_{eq})^{(n-k)}$. We then cause the state to ‘isothermally expand’ in each degenerate subspace, taking the state from the initial pure state to the final maximally mixed state and in the process extracting $T \log d$ energy from the bath, where d is the dimension of the subspace. In total, the final average energy of the ladder will be

$$\langle E \rangle_w = \sum_{k=0}^n \binom{n}{k} p_{eq}^k (1-p_{eq})^{(n-k)} T \log \binom{n}{k} \quad (F2)$$

In the asymptotic limit, as $n \rightarrow \infty$ it is straightforward to show that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^n \binom{n}{k} p_{eq}^k (1-p_{eq})^{(n-k)} T \log \binom{n}{k} = TH(p_{eq}) \quad (F3)$$

since in the limit the only contribution to the sum comes from $k = np_{eq}$. We thus see that asymptotically the expected amount of work is extracted. There are a couple of comments which are worth mentioning. First we note that due to concentration effects, we know that almost all weight lies in those subspaces with energy $np_{eq} \pm \mathcal{O}(\sqrt{n})$. We could therefore ignore those subspaces which lie outside this range and obtain, to leading order, the same amount of work. The results obtain however are independent of the concentration effects. Second we recall that the protocol starts with collective-dephasing in the energy eigenbasis. It might appear that this operation, which changes the entropy of the system, throws away work. However, asymptotically we obtain to first order the optimal amount of work, and so we conclude that the amount thrown away must scale sub-linearly in n . In fact the amount of entropy thrown away is only $\mathcal{O}(\log n)$, the entropy of the binomial distribution, since the dephasing leaves the system binomially distributed amongst energy eigenspaces which have orthogonal support. This shows that asymptotically the coherence between states of different energy are thermodynamically insignificant.

2. General work extraction protocol

Finally, for the case of arbitrary states of arbitrary dimension d , we note that the same conclusion can also be drawn. First we need to show that the operation of collective-dephasing never increases the entropy of a collection of systems too much in the asymptotic limit. Let us denote the projectors onto the energy eigenspaces by Π_k , and the total number of different energies by K . After the operation of collective-dephasing the state becomes block diagonal in the energy eigenbasis, namely

$$\rho^{\otimes n} \mapsto \rho' = \sum_{k=0}^{K-1} \Pi_k \rho^{\otimes n} \Pi_k \quad (F4)$$

We are interested in the difference in entropy $S(\rho^{\otimes n}) - S(\rho')$, i.e. the entropy generated by this operation. To do so, let us first consider the coherent version of collective-dephasing. Let us introduce an ancillary system A with K states $|z\rangle_A$ prepared initially in the state $|0\rangle_A$. Let us also consider the unitary operation V

$$V = \sum_{k,z} \Pi_k \otimes |z+k \bmod K\rangle_A \langle z| \quad (F5)$$

Applied to the state $\rho_{SA} = \rho^{\otimes n} \otimes |0\rangle_A \langle 0|$ this performs the transformation

$$V(\rho^{\otimes n} \otimes |0\rangle_A \langle 0|) V^\dagger = \rho'_{SA} = \sum_{k,k'} \Pi_k \rho^{\otimes n} \Pi_{k'} \otimes |k\rangle_A \langle k'| \quad (F6)$$

It is straightforward to see that $\text{tr}_A \rho'_{SA} = \rho'$ whilst

$$\text{tr}_S \rho'_{SA} \equiv \rho'_A = \sum_k \text{tr}(\Pi_k \rho^{\otimes n}) |k\rangle_A \langle k| \quad (F7)$$

Considering now the entropies, we have $S(\rho'_{SA}) = S(\rho^{\otimes n}) = nS(\rho)$, since the two states differ by the appended pure ancilla and the joint unitary transformation. Thus, from the Araki-Lieb inequality $S(\rho_{SA}) \geq |S(\rho_S) - S(\rho_A)|$ we have

$$nS(\rho) \geq |S(\rho') - S(\rho'_A)| \quad (F8)$$

Next, note that we can rewrite ρ' in the form

$$\rho' = \sum_k \text{tr}(\Pi_k \rho^{\otimes n}) \frac{\Pi_k \rho^{\otimes n} \Pi_k}{\text{tr}(\Pi_k \rho^{\otimes n})} \quad (F9)$$

where the sum is restricted to those k for which $\text{tr}(\Pi_k \rho^{\otimes n}) \neq 0$. Comparing this to (F7), we note that $S(\rho') \geq S(\rho'_A)$, as a mixture of orthogonal mixed states always has higher entropy than the same mixture of orthogonal pure states. Hence the absolute value can be omitted on the right-hand-side of equation (F8) and after rearranging we get

$$S(\rho'_A) \geq S(\rho') - nS(\rho) \quad (F10)$$

Next, note that $S(\rho'_A) \leq \log K$, since in the worst case ρ'_A is maximally mixed over K states. Hence we find

$$\log K \geq S(\rho') - nS(\rho) \quad (\text{F11})$$

The number of different energy levels K is upper bounded by the number of ways of distributing n indistinguishable particles between d distinguishable energy levels, which is given by $\binom{n+d-1}{n}$. Thus

$$S(\rho') - nS(\rho) \leq \log \binom{n+d-1}{n} \leq (d-1) \log(n+1) \quad (\text{F12})$$

This demonstrates that the operation of collective-dephasing on n qubits can only increase the entropy by a logarithmic amount in n . The increase *per qubit* is thus of the order $\frac{d-1}{n} \log(n+1)$, which vanishes in the asymptotic limit. This result does not follow from any concentration of measure, but simply by the fact that the number of energy eigenvalues of n non-interacting systems is logarithmic in n .

Appendix G: Discrete work storage

In this appendix we will show that the energy level spacing of the weight need only become continuous in the asymptotic limit and that for any finite number of bath qubits N it is sufficient to consider the levels to have a minimum spacing \mathcal{E} which scales as $1/N^2$.

Let us begin by recalling how the protocol works. We pick a collection of N bath qubits which have excited state probabilities $q^{(k)}$ given by

$$q^{(k)} = p + \frac{k}{N}(p_{eq} - p) \quad (\text{G1})$$

and therefore excited state energies $E_B^{(k)} = TS'_c(q^{(k)})$. At each stage of the protocol the weight either gains or loses energy $E_S - E_B^{(k)}$.

Let us now assume instead that the of being a continuous degree of freedom that the weight has a discrete (but infinite) set of states $|k\rangle_w$ and Hamiltonian

$$H_w = \sum_{k=-\infty}^{\infty} k \mathcal{E} |k\rangle_w \langle k| \quad (\text{G2})$$

Accordingly, we shall adjust the energy levels of the bath qubits to

$$\tilde{E}_B^{(k)} = E_B^{(k)} + \varepsilon^{(k)} \quad (\text{G3})$$

such that

$$E_S - \tilde{E}_B^{(k)} = m^{(k)} \mathcal{E} \quad (\text{G4})$$

where the $m^{(k)}$ are integers, and we take the smallest possible $\varepsilon^{(k)}$ possible such that this condition is met, meaning that $|\varepsilon^{(k)}| < \mathcal{E}$.

To proceed, we first note that (D1) can alternatively be re-written as

$$\begin{aligned} \langle E \rangle_w^{[0]} &= -p_{eq} E_S + p E_B^{(1)} + \sum_{k=1}^N (q^{(k)} - q^{(k-1)}) E_B^{(k)} \\ &= \sum_{k=0}^{N-1} q^{(k)} (E_B^{(k)} - E_B^{(k+1)}) \end{aligned} \quad (\text{G5})$$

Thus upon making the replacement $E_B^{(k)} \rightarrow \tilde{E}_B^{(k)}$ for the case of a discrete weight, the average energy, conditional on the qubit starting in the ground state becomes

$$\begin{aligned} \langle \tilde{E} \rangle_w^{[0]} &= \sum_{k=0}^{N-1} q^{(k)} (\tilde{E}_B^{(k)} - \tilde{E}_B^{(k+1)}) \\ &= \sum_{k=0}^{N-1} q^{(k)} (E_B^{(k)} - E_B^{(k+1)}) + \sum_{k=0}^{N-1} q^{(k)} (\varepsilon^{(k)} - \varepsilon^{(k+1)}) \\ &= \langle E \rangle_w^{[0]} + \sum_{k=0}^{N-1} q^{(k)} (\varepsilon^{(k)} - \varepsilon^{(k+1)}) \end{aligned} \quad (\text{G6})$$

It is the final term in the last line which is the error introduced by considering a discrete weight, and thus it is the magnitude of this term which is of interest. This can easily be bounded as follows

$$\begin{aligned} \left| \sum_{k=0}^{N-1} q^{(k)} (\varepsilon^{(k)} - \varepsilon^{(k+1)}) \right| &\leq \sum_{k=0}^{N-1} q^{(k)} |\varepsilon^{(k)} - \varepsilon^{(k+1)}| \\ &\leq \sum_{k=0}^{N-1} q^{(k)} 2\mathcal{E} \\ &= \mathcal{E} ((p + p_{eq})N + (p - p_{eq})) \end{aligned} \quad (\text{G7})$$

Therefore as long as the product $\mathcal{E}N \rightarrow 0$ as $N \rightarrow \infty$, then this term will asymptotically vanish. Choosing $\mathcal{E} = \epsilon/N^2$ would ensure that this correction is $\mathcal{O}(1/N)$ and thus the same order as the leading order correction to $\langle E \rangle_w^{[0]}$ for finite N . An identical analysis holds also for $\langle E \rangle_w^{[1]}$, the average energy of the weight conditional on the system starting in the state $|1\rangle_s$, and therefore we conclude that the results more generally do not rely crucially on the continuous nature of the weight.